

MPEG-4 Speech coding

Masayuki Nishiguchi

**Audio and Speech Group, HomeNet Processing Lab,
HomeNet Laboratories, Sony Corporation**

Outline

Background

MPEG-4 Speech coding - features

 MPEG-4 CELP

 Algorithm

 Performance

 Demonstration

 MPEG-4 HVXC

 Algorithm

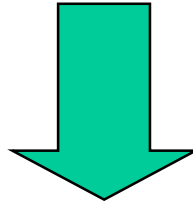
 Performance

 Demonstration

Summary

Background

Most of the existing speech coding standards support only a single “compression” functionality.

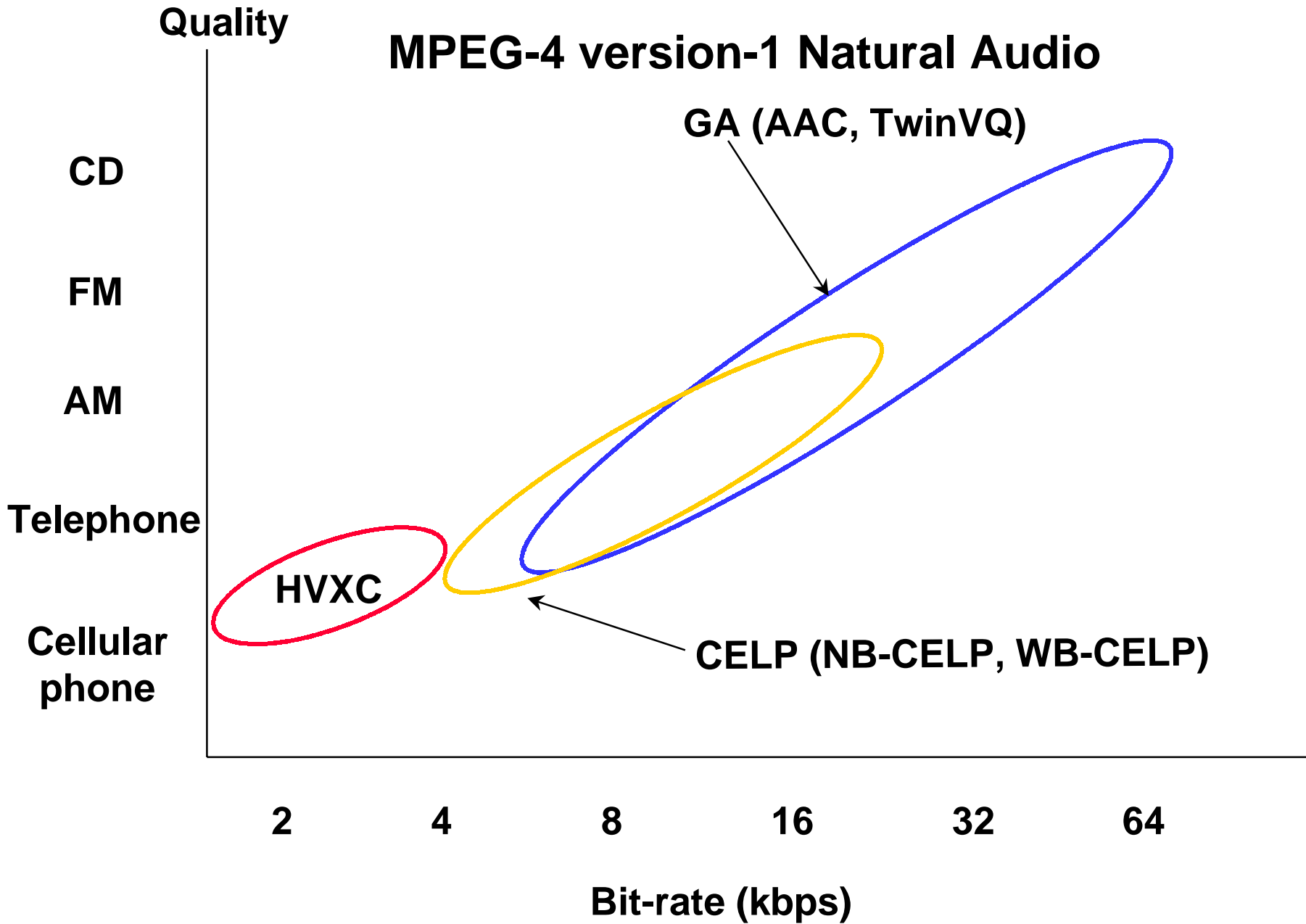


Speech coding algorithms with **high coding efficiency and multiple functionalities** play important role for efficient use of bandwidth and emerging new applications of multimedia systems.

MPEG-4 Speech Coding - features

- Two basic algorithms
 - HVXC (Harmonic Vector eXcitation Coding)
 - CELP (Code Excited Linear Prediction)
- Multi bit-rates
 - 1.5 ~ 24 kbps
- Narrow-band and wide-band - CELP
- Lowest bit-rate as an international standard coding - HVXC
 - 2.0 kbps (fixed) ave 1.5 kbps (var)
- New Functionalities
 - Speed / Pitch change - HVXC
 - Bit-rate scalability - HVXC, CELP
 - Bandwidth scalability - CELP

MPEG-4 version-1 Natural Audio



MPEG-4 CELP

Narrow band 3.85-12.2 kbps 10-40 ms frame

Wide band 10.9-23.8 kbps 10-20 ms frame

Multi-rate 200 - 800 bps step

Bit-rate scalability - 2.0kbps(NB), 4.0kbps(WB) step

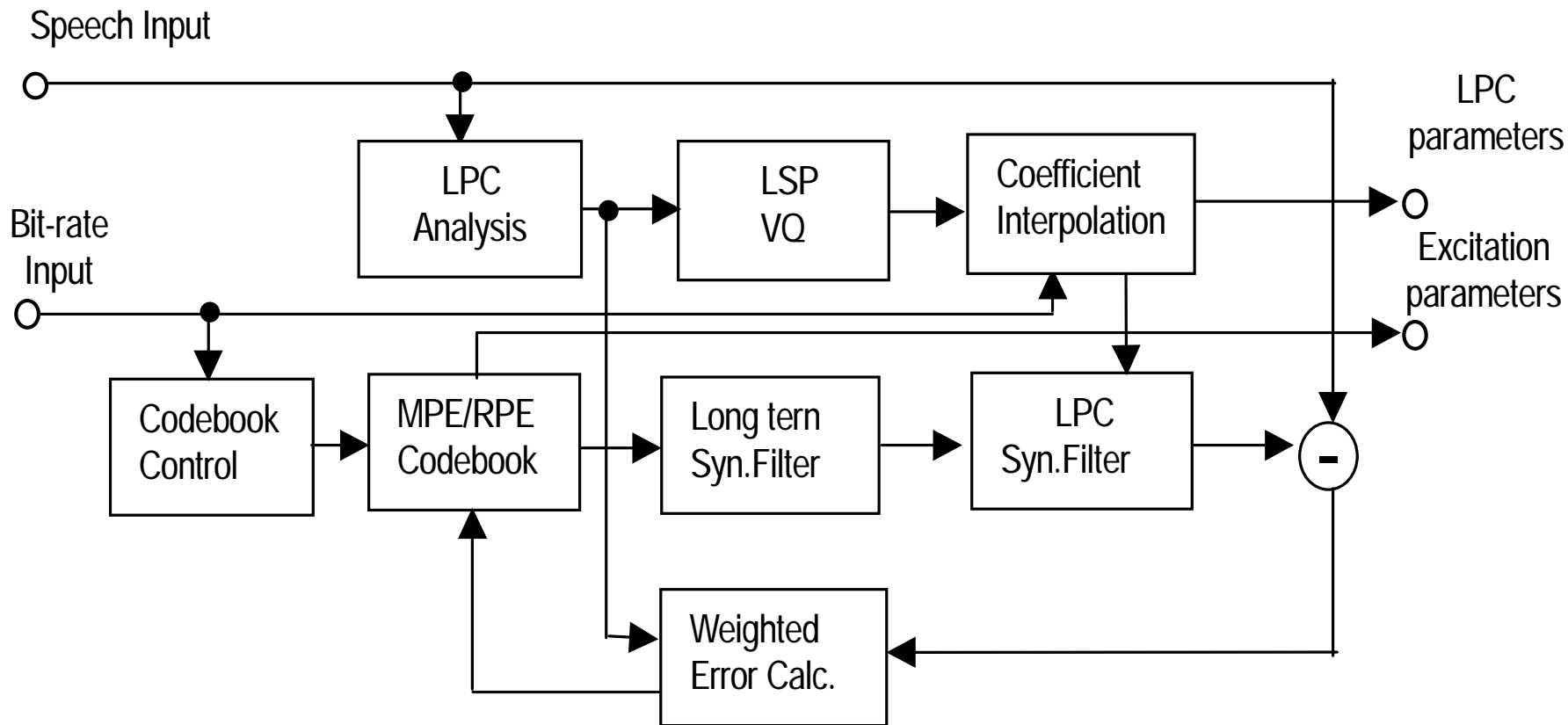
Bandwidth scalability

Fine rate control

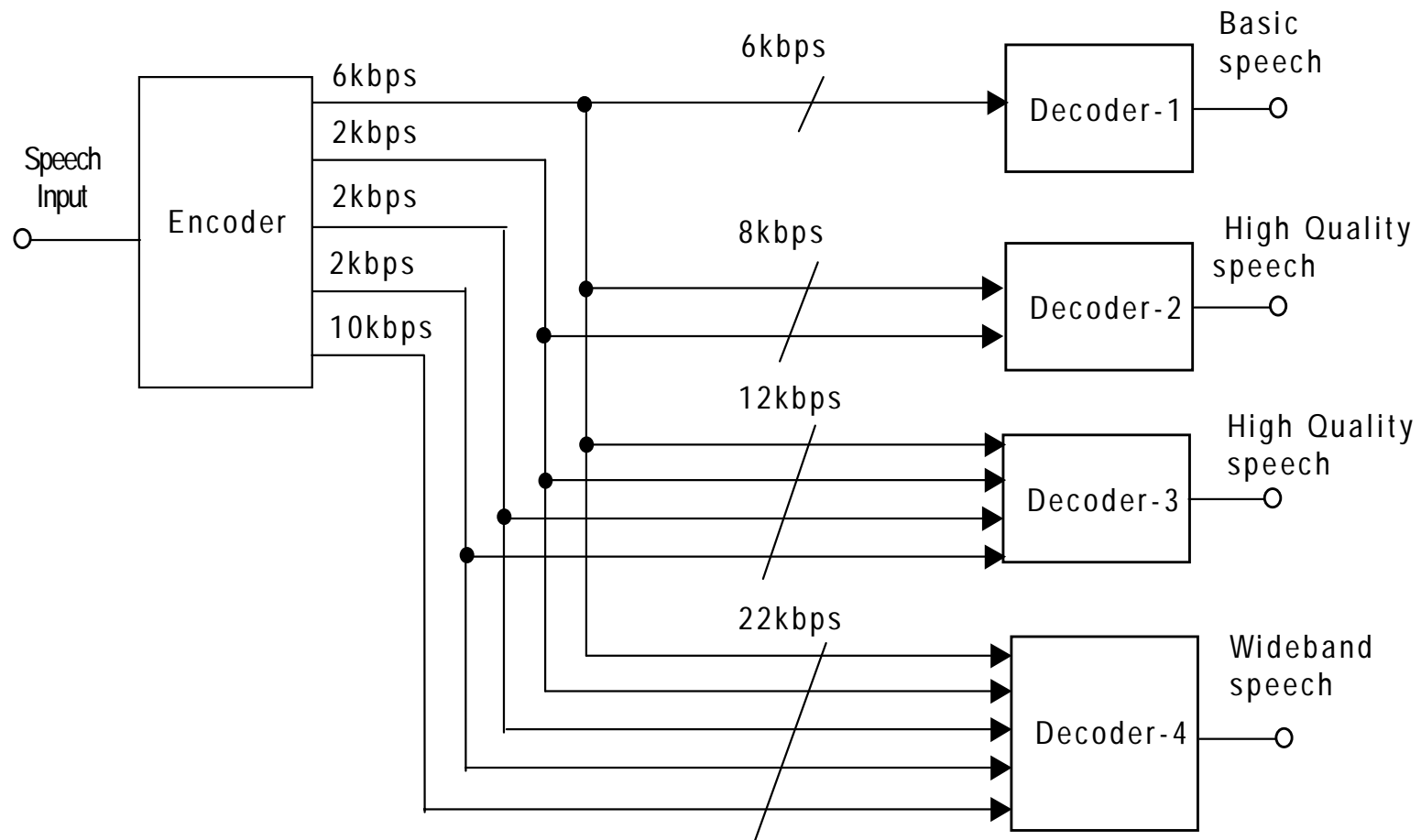
Regular pulse - WB: Low complexity

Multi pulse - WB, NB: High coding efficiency

Blockdiagram of the CELP encoder



Structure of the bit-rate scalable coding



Performance

Speech quality of MPEG-4 speech codecs were evaluated in the official MPEG-4 verification tests in Aug 98 at 2 European labs and 1 Japanese lab*.

15 Japanese items were evaluated by 16 Japanese listeners in the Japanese Lab.

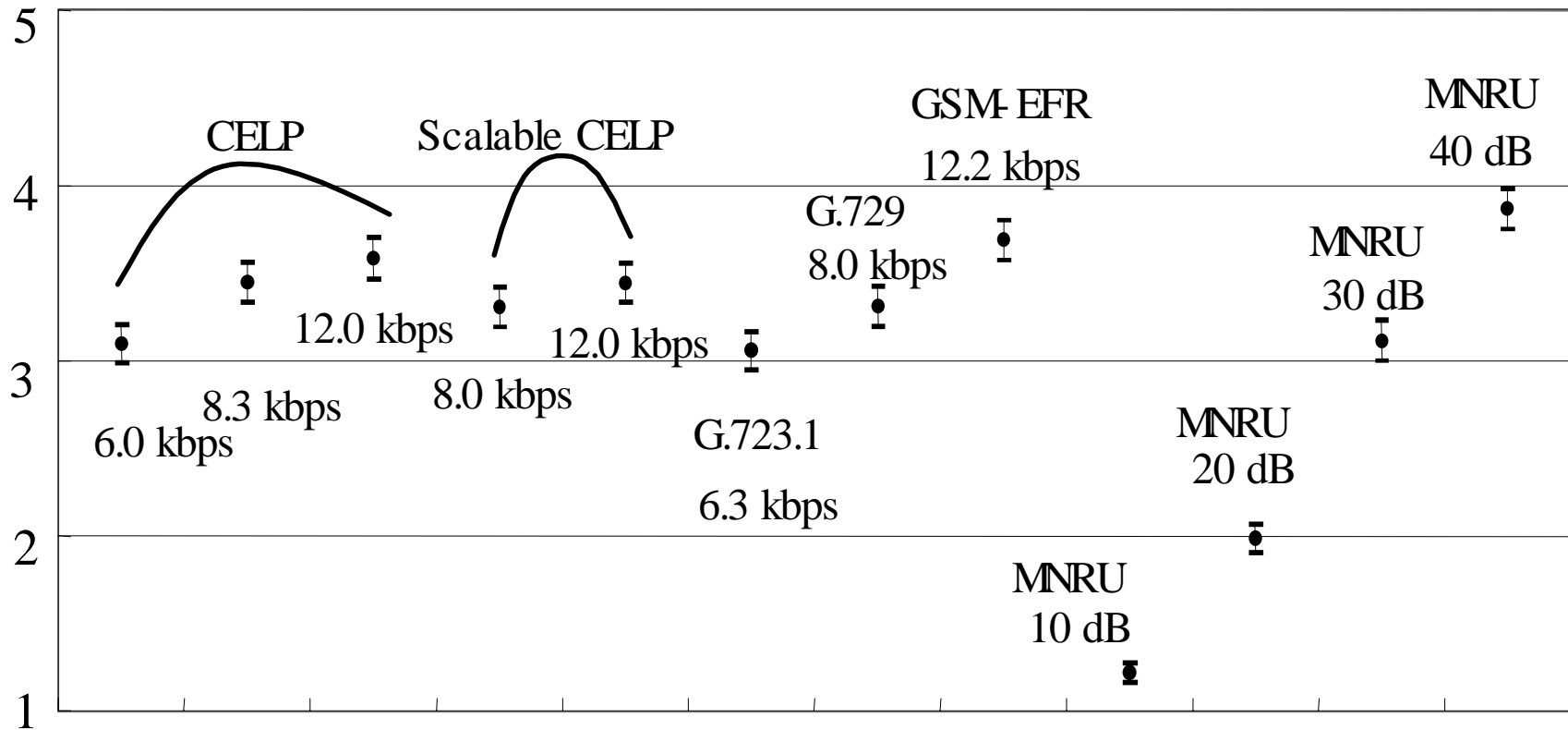
15 English/German/Swedish items were evaluated by 18 German and 16 Finnish listeners in the European Labs.

* ISO/IEC JTC1/SC29/WG11 MPEG98/N2424 "Report on the MPEG-4 speech codec verification tests," Oct. 1998

Performance

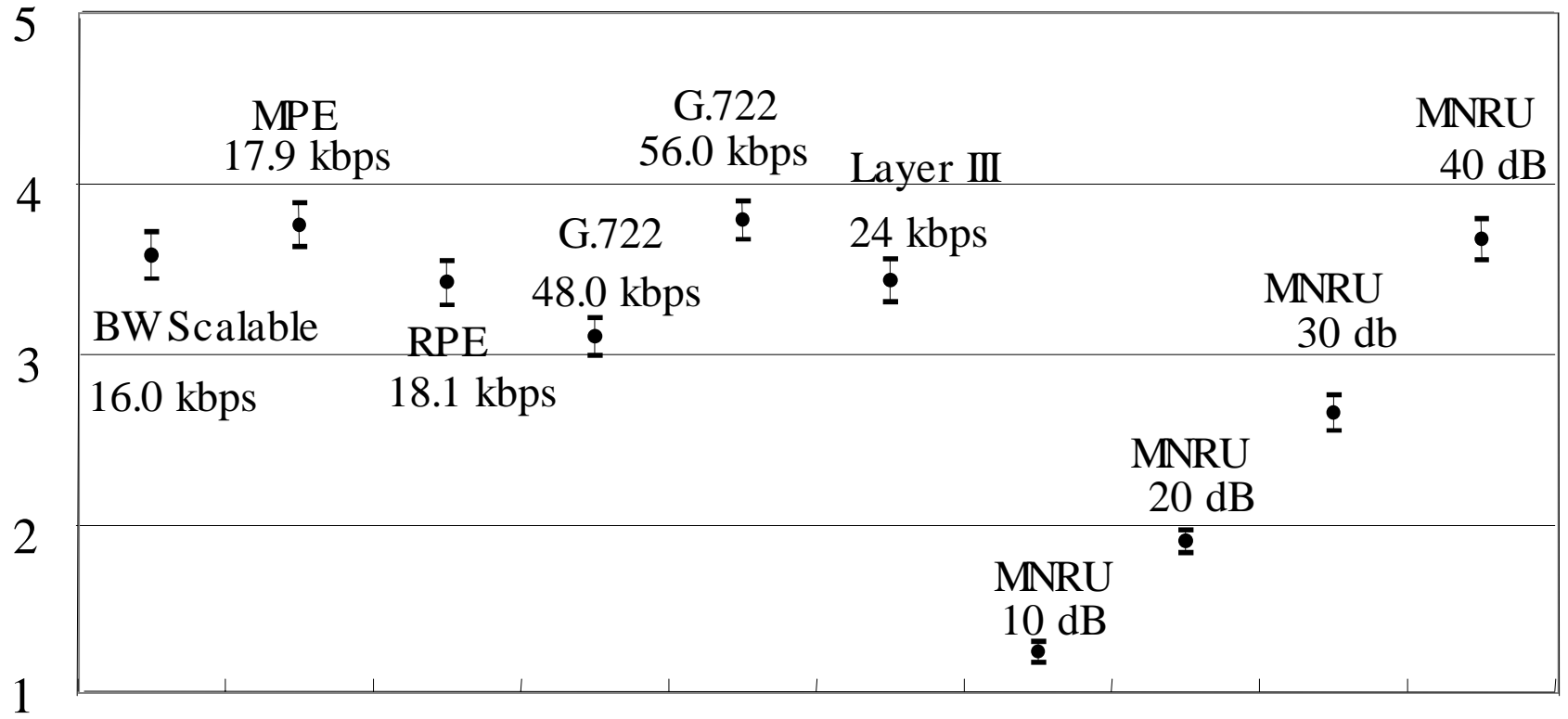
MOS 95%CI

Narrow band CELP - Japanese



MOS 95%CI

Wideband CELP - Japanese



ISO/IEC JTC1/SC29/WG11 MPEG98/N2424 "Report on the MPEG-4 speech codec verification tests," Oct. 1998

MPEG-4 CELP demonstration

6 kbps NB CELP

12 kbps NB CELP

22 kbps WB CELP
(BW-scalable)

CELP Demo samples are generated by NEC

MPEG-4 HVXC

Low bit-rate / good quality

- 2.0 / 4.0kbps (fixed) , 1.5 / 3.0kbps ave. (variable)
- HVXC at 2.0kbps has higher quality than FS1016 CELP at 4.8kbps

Bit-rate scalability

- 2.0kbps decoding is possible using 4.0kbps bit-stream

Speed change & Pitch change

- Attractive for fast speech database search & browsing

Approach

Two different types of coding schemes are combined. One is suitable for voiced segments and the other for unvoiced segments.

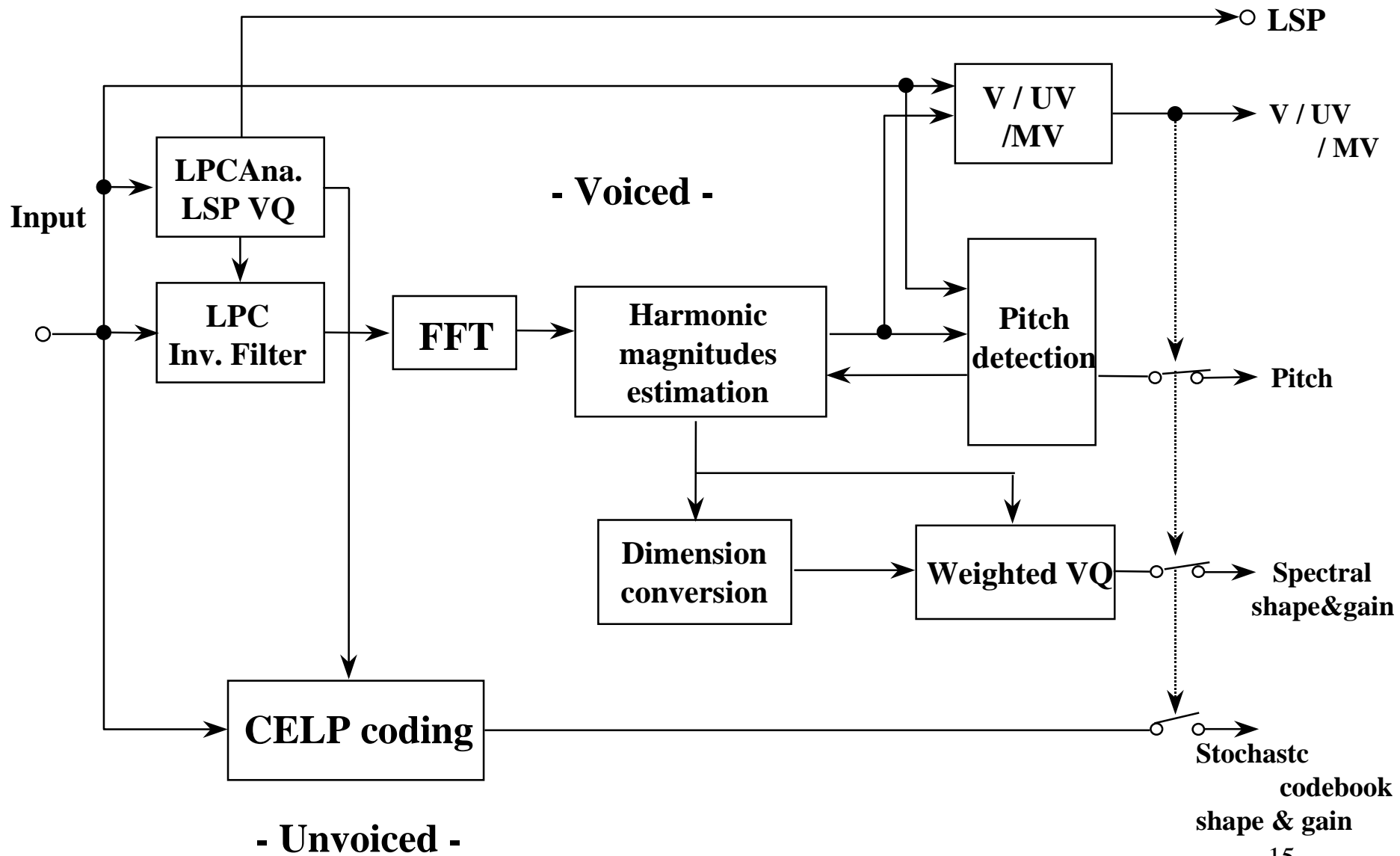
Voiced: Phase information is thrown away by harmonic representation of power spectrum of LPC residual.

Frequency domain analysis / synthesis.

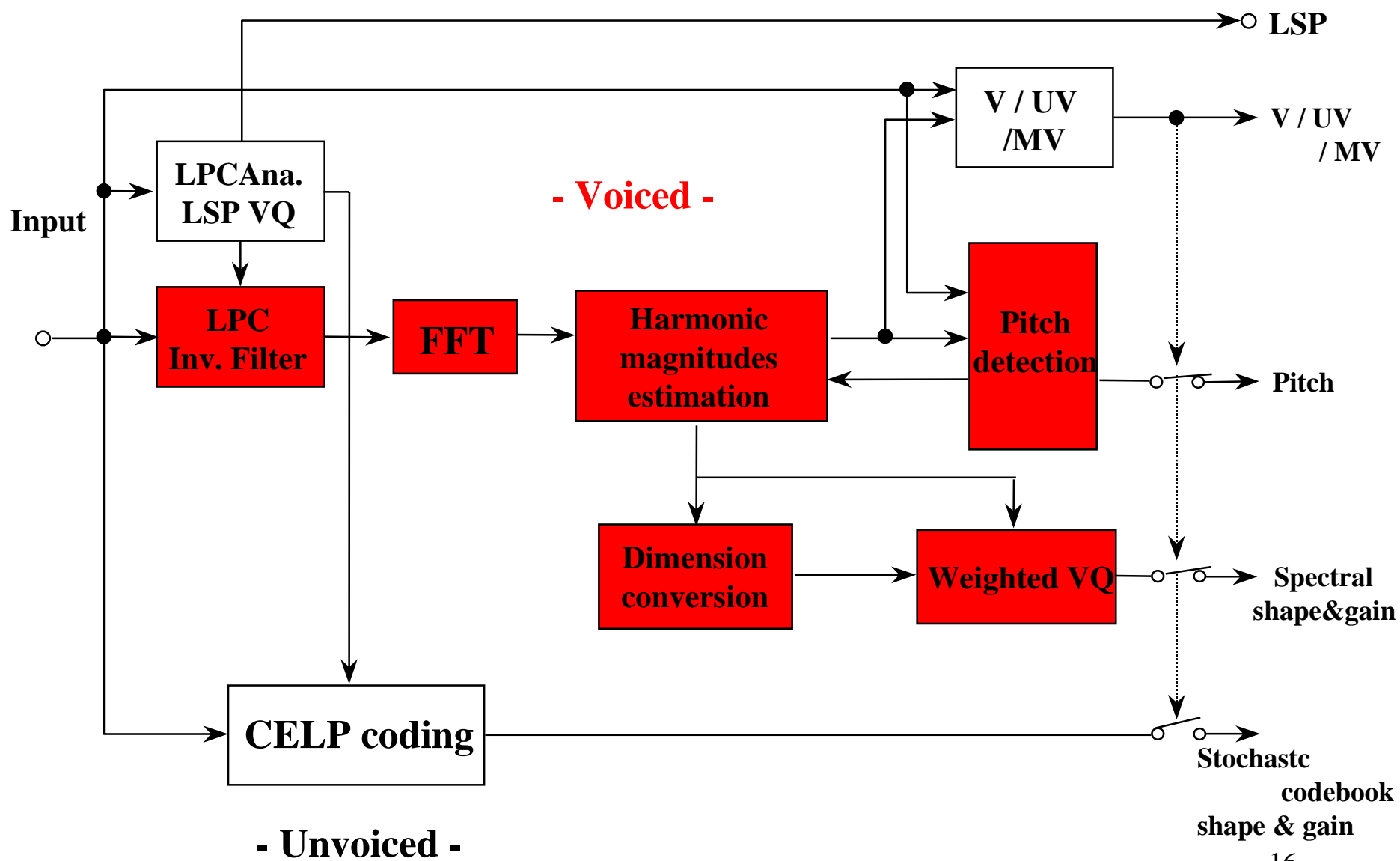
Unvoiced: Crisp consonant is obtained by CELP coding.

Time domain analysis / synthesis.

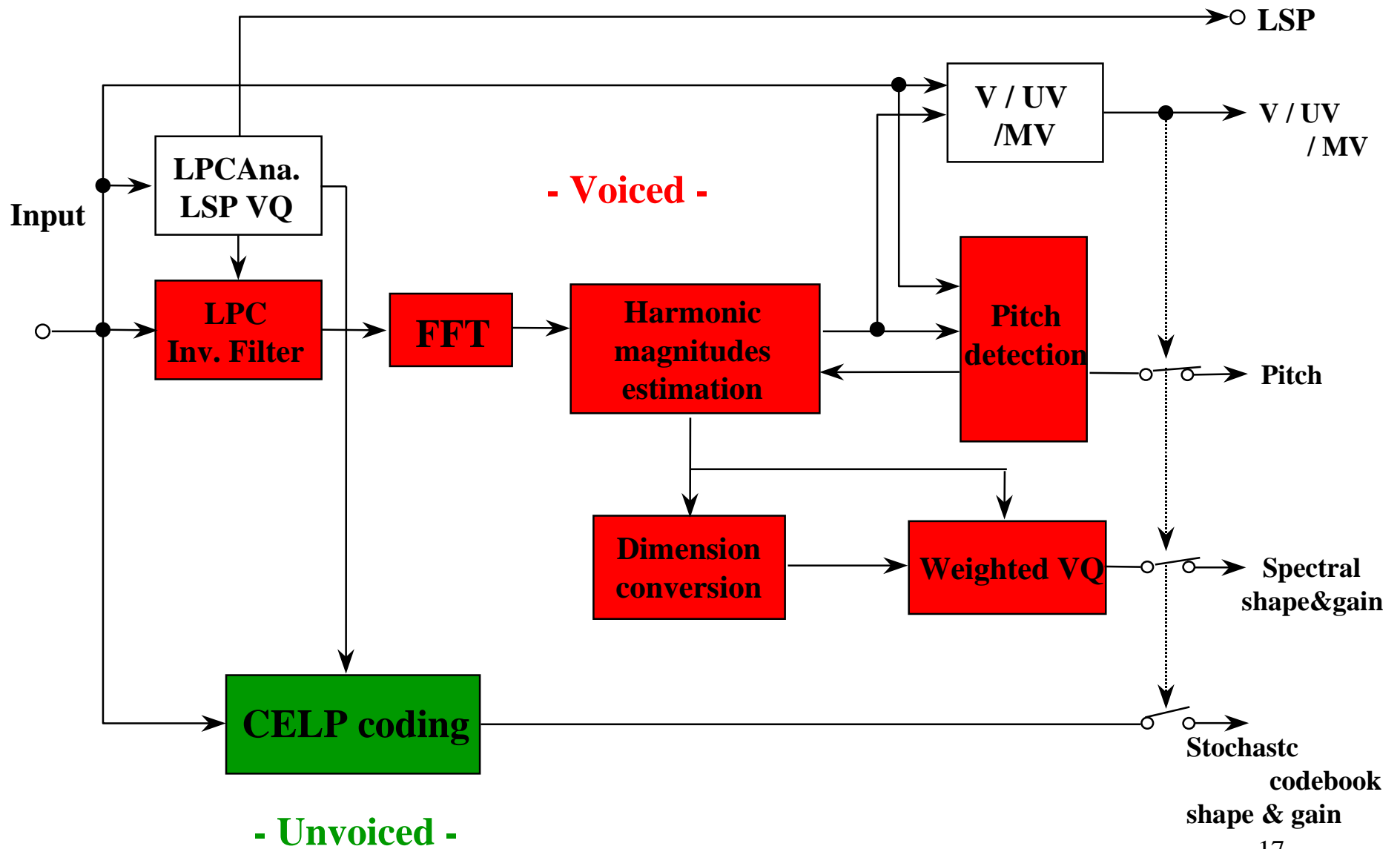
Encoder



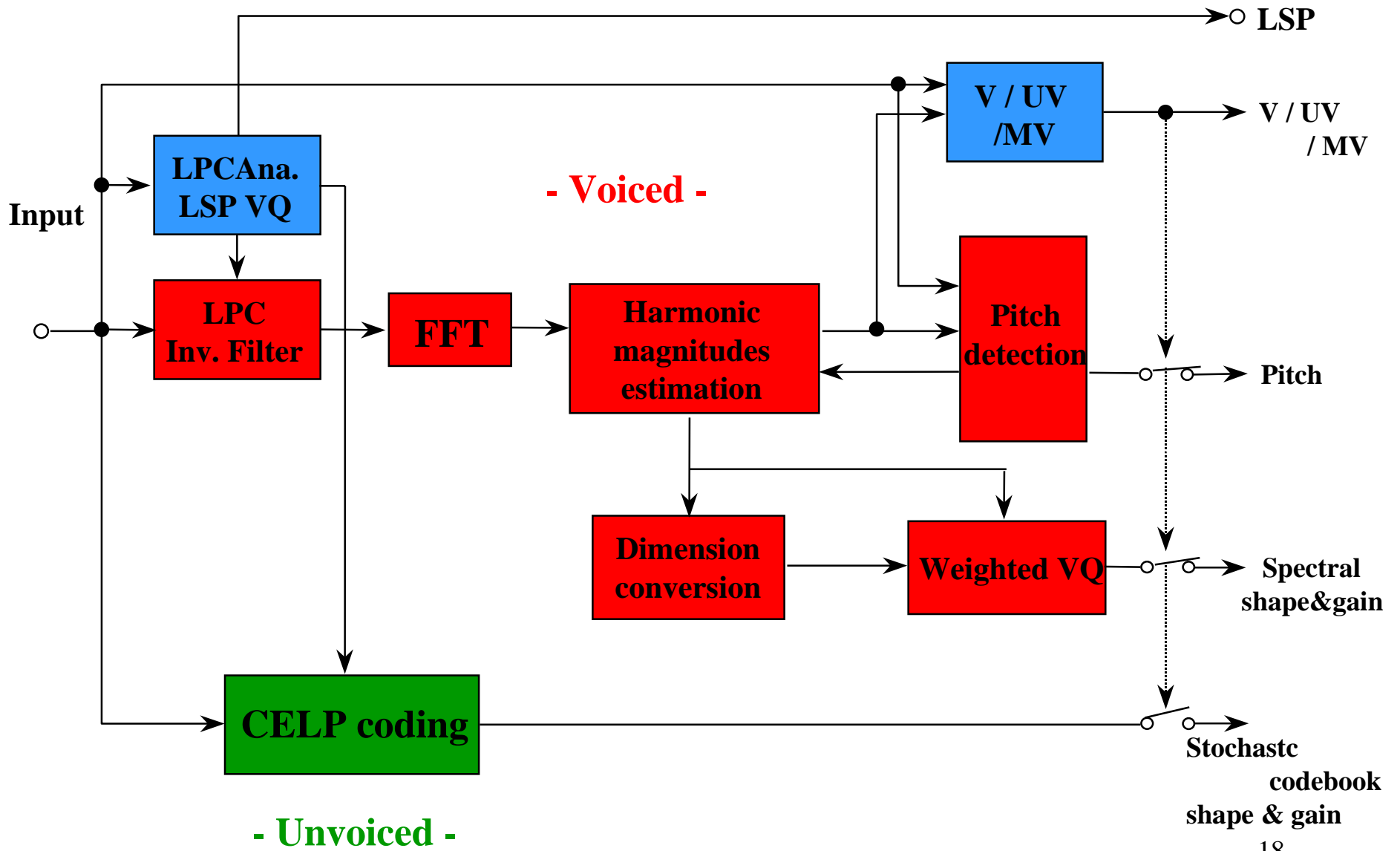
Encoder



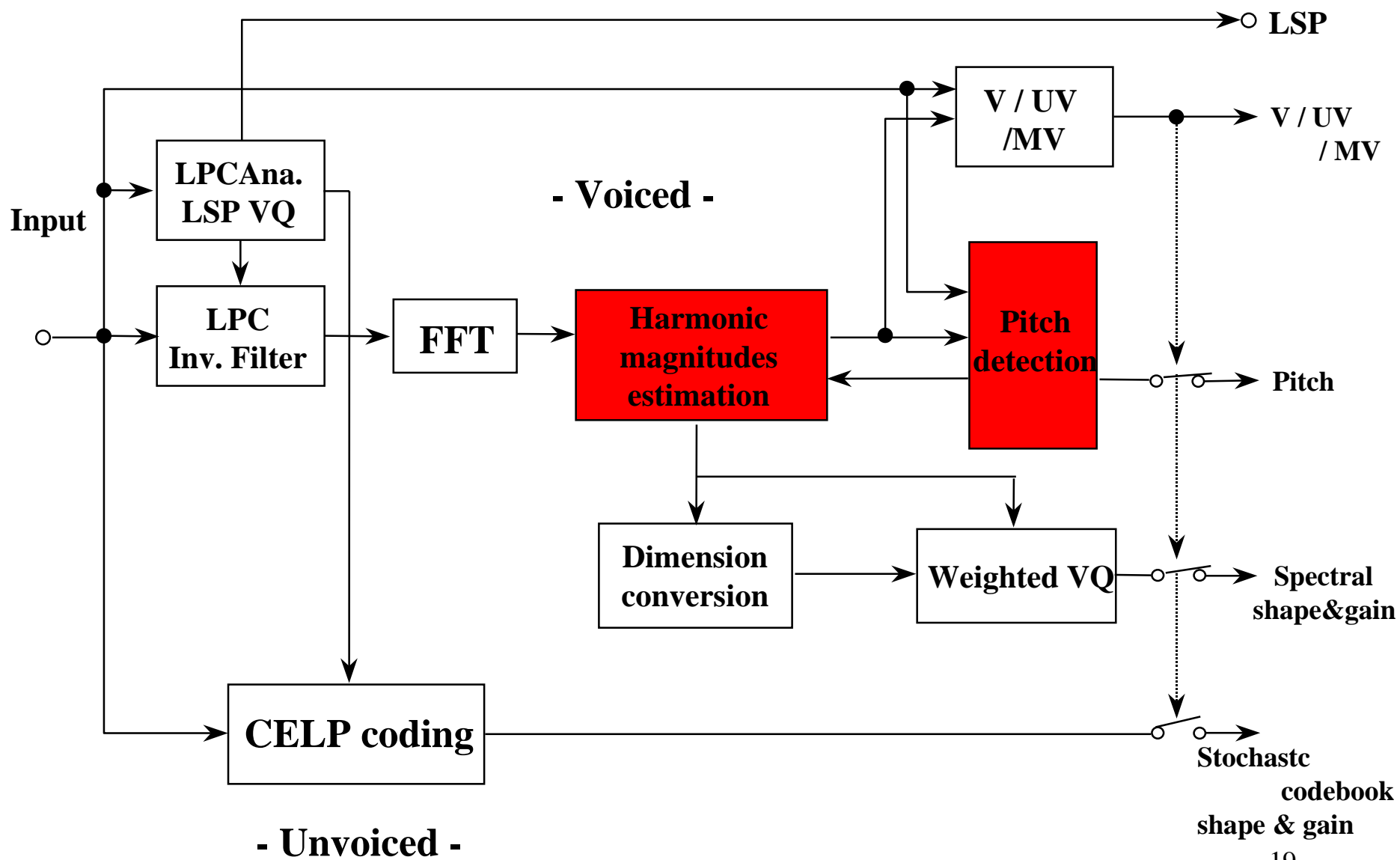
Encoder



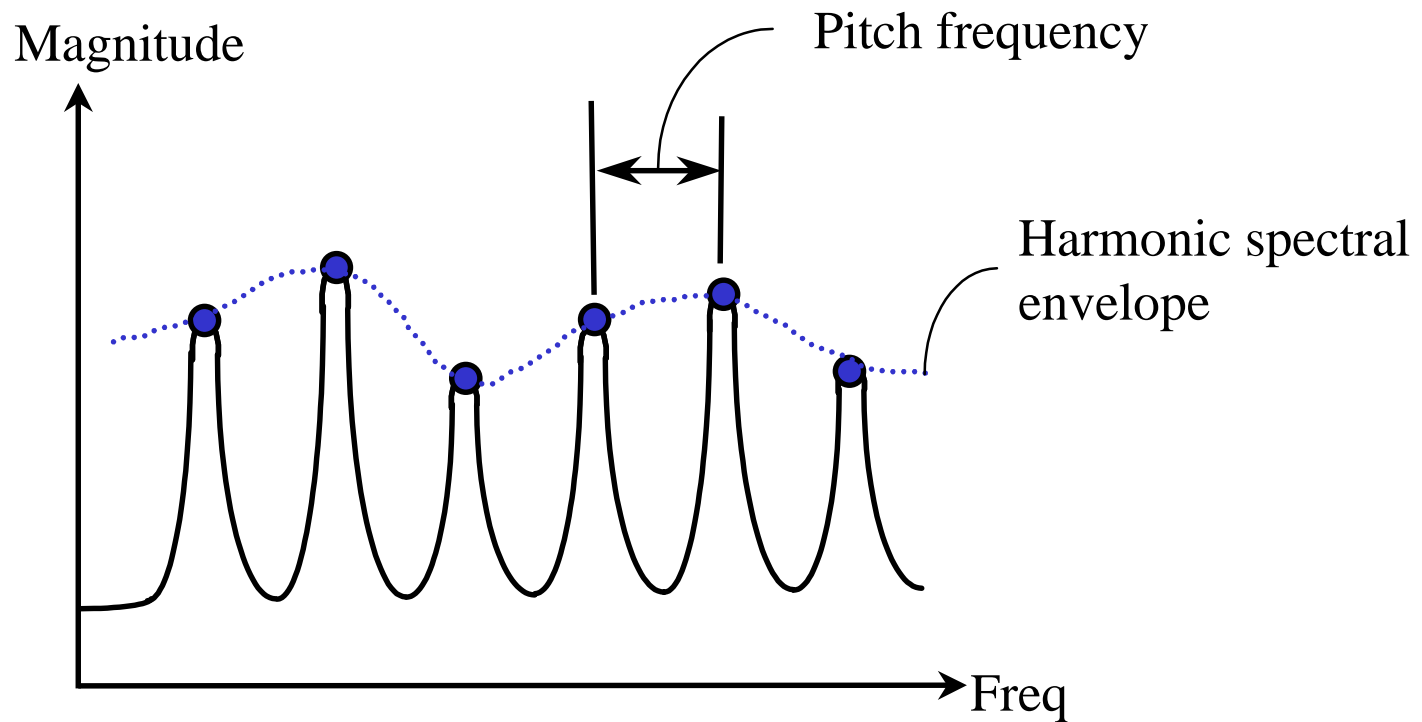
Encoder



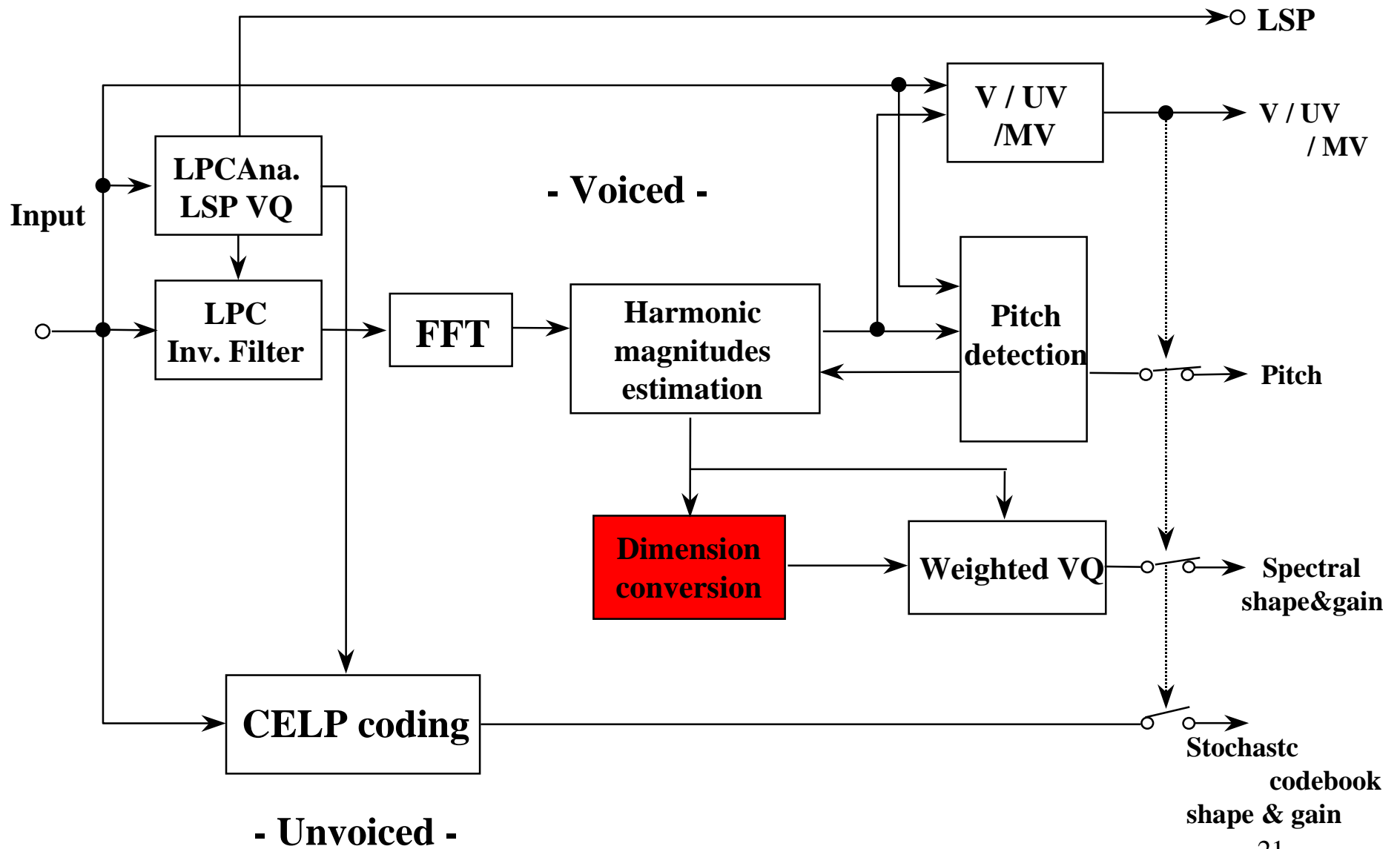
Encoder



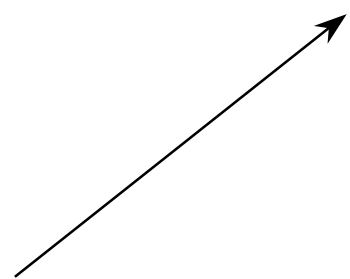
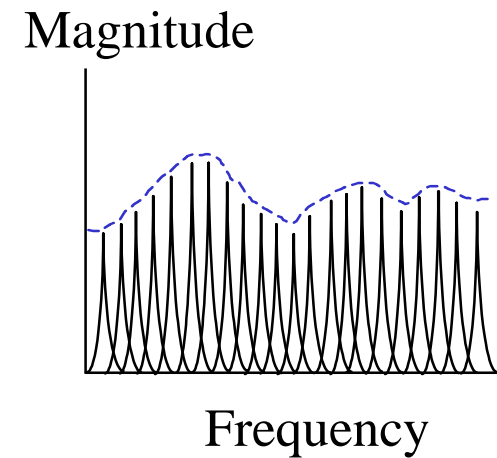
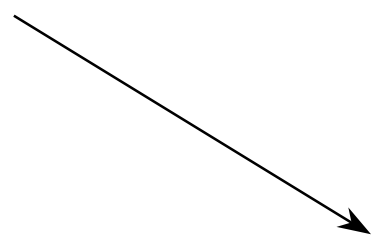
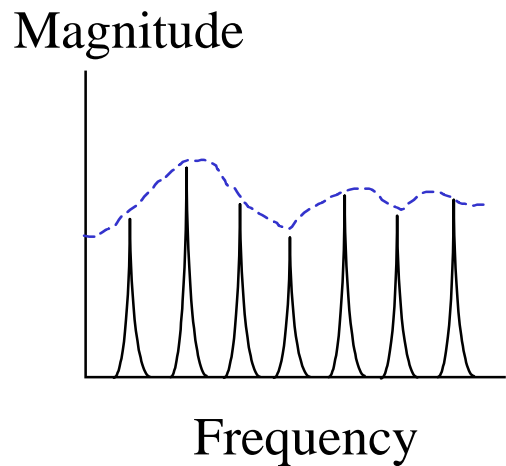
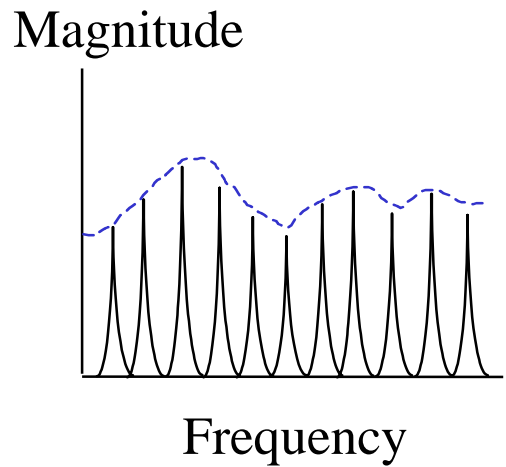
Harmonic spectral magnitudes and fine pitch estimation



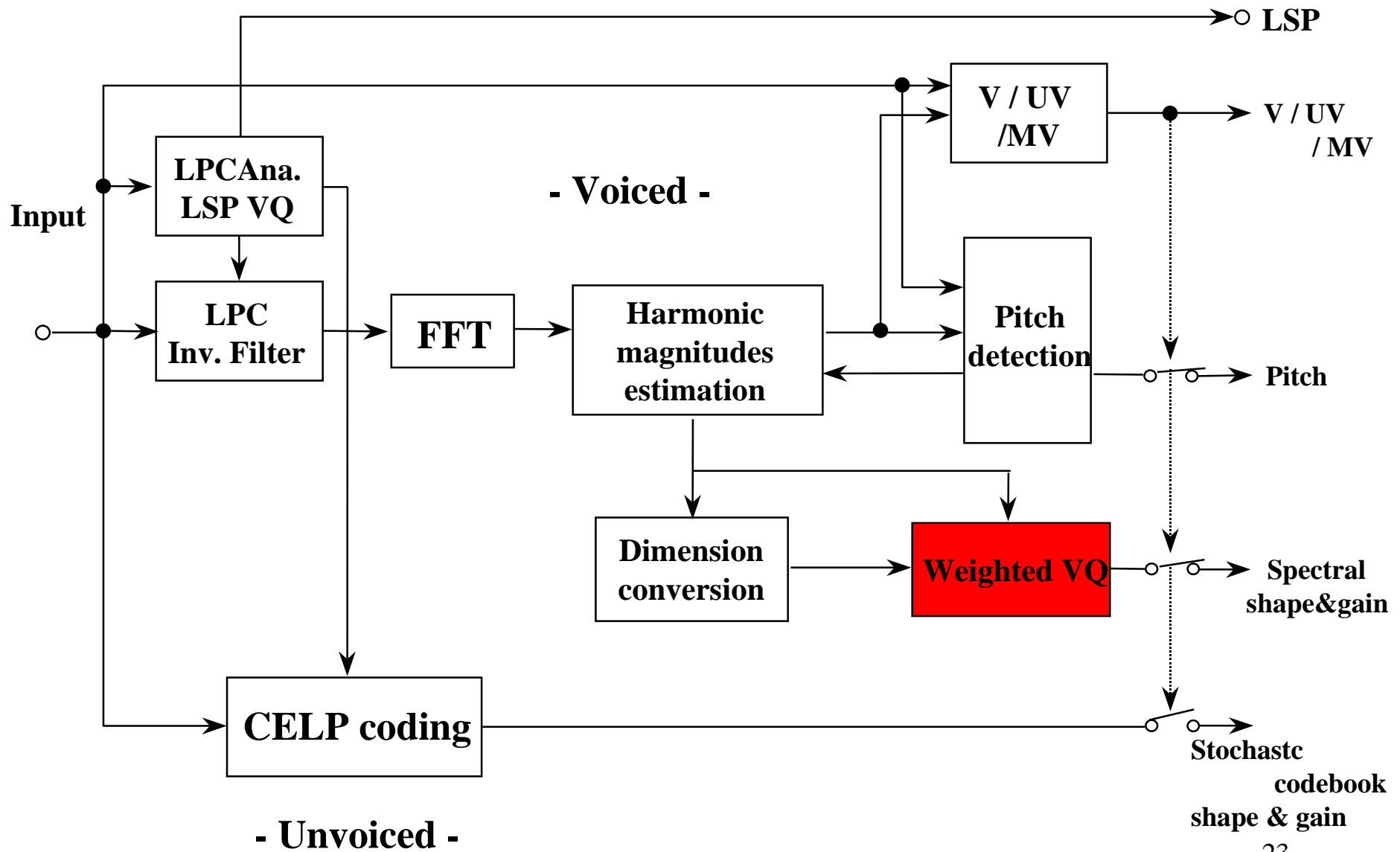
Encoder



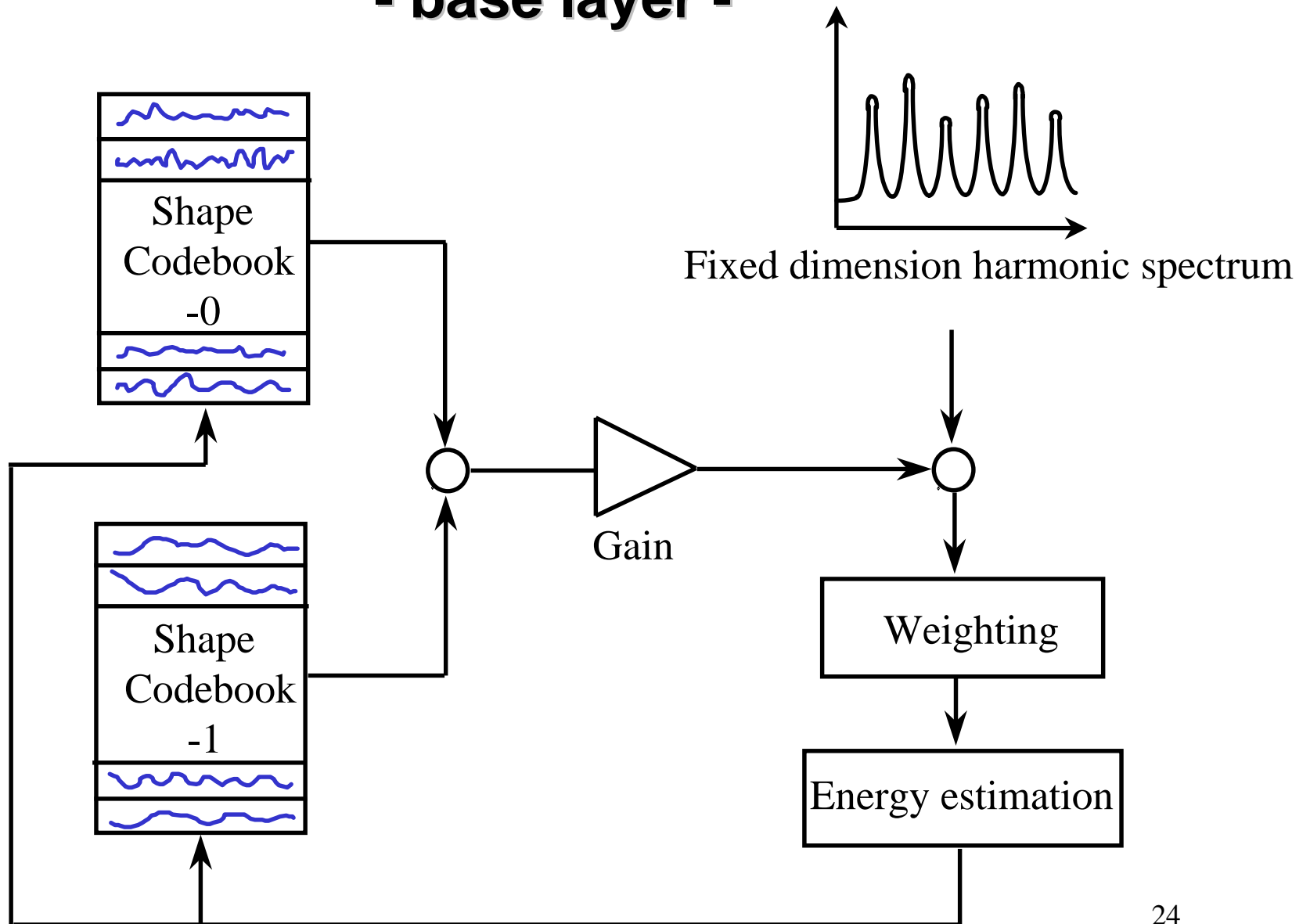
Dimension conversion of Harmonic spectral magnitudes



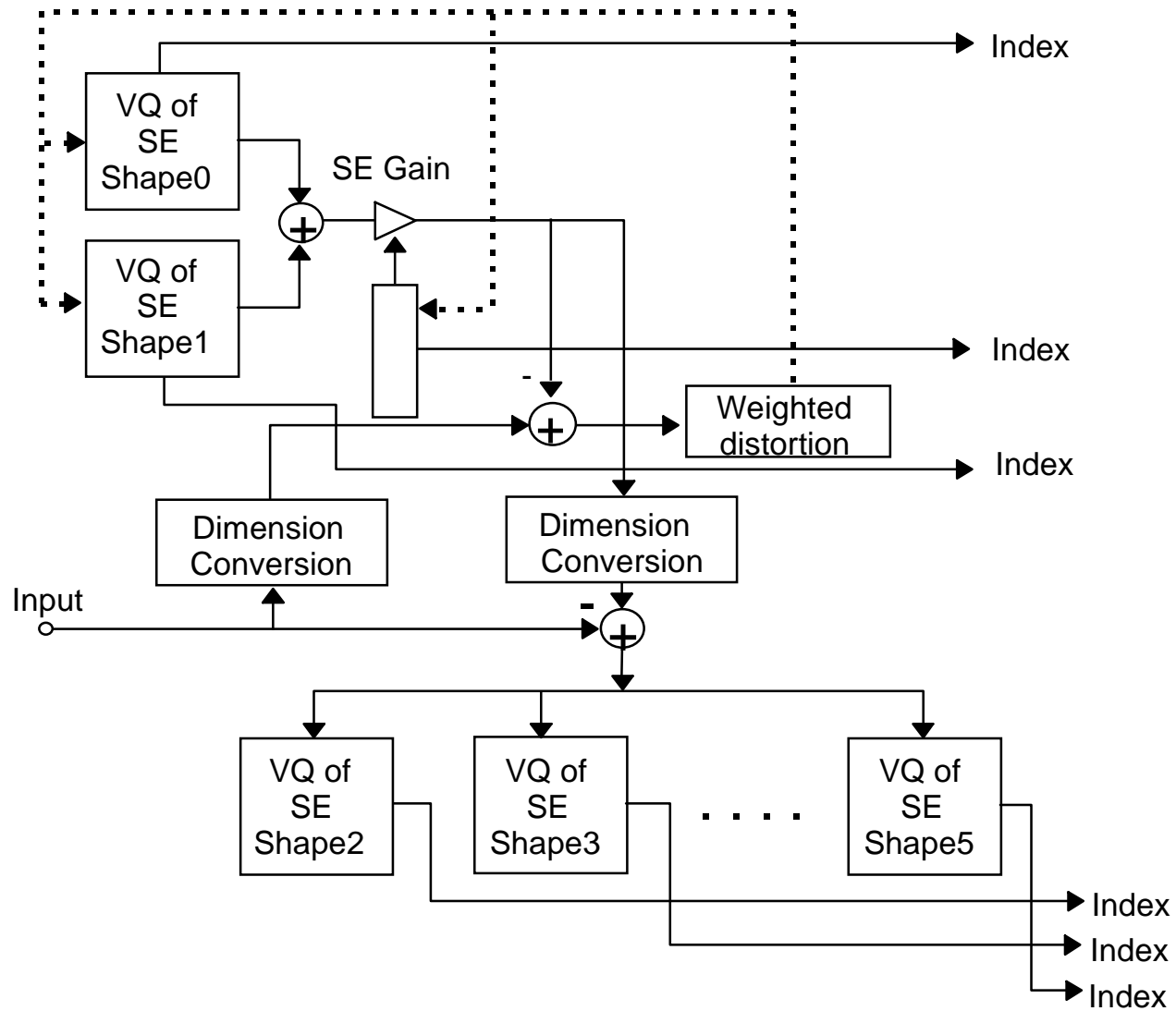
Encoder



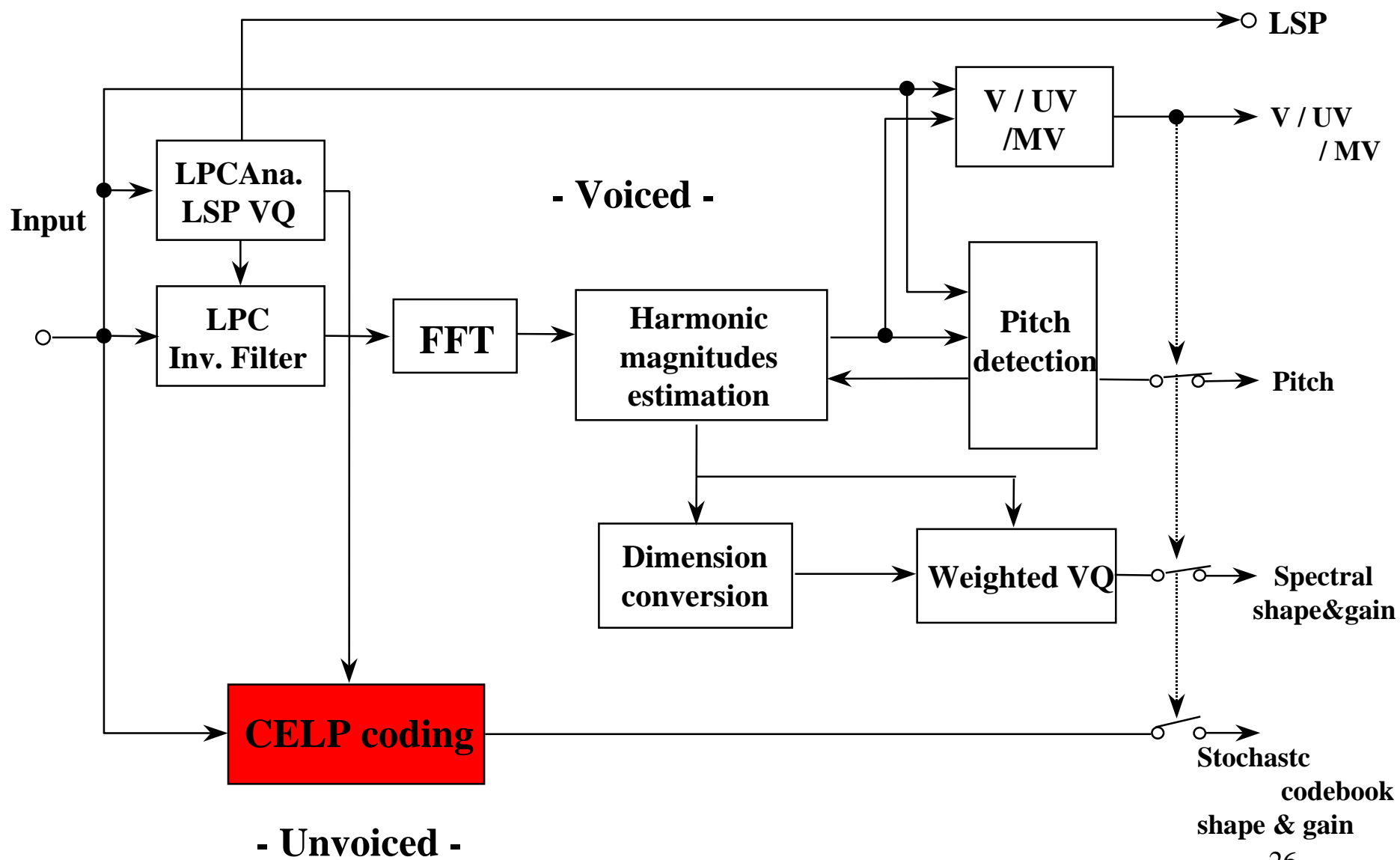
Vector quantization of harmonic spectral envelope - base layer -



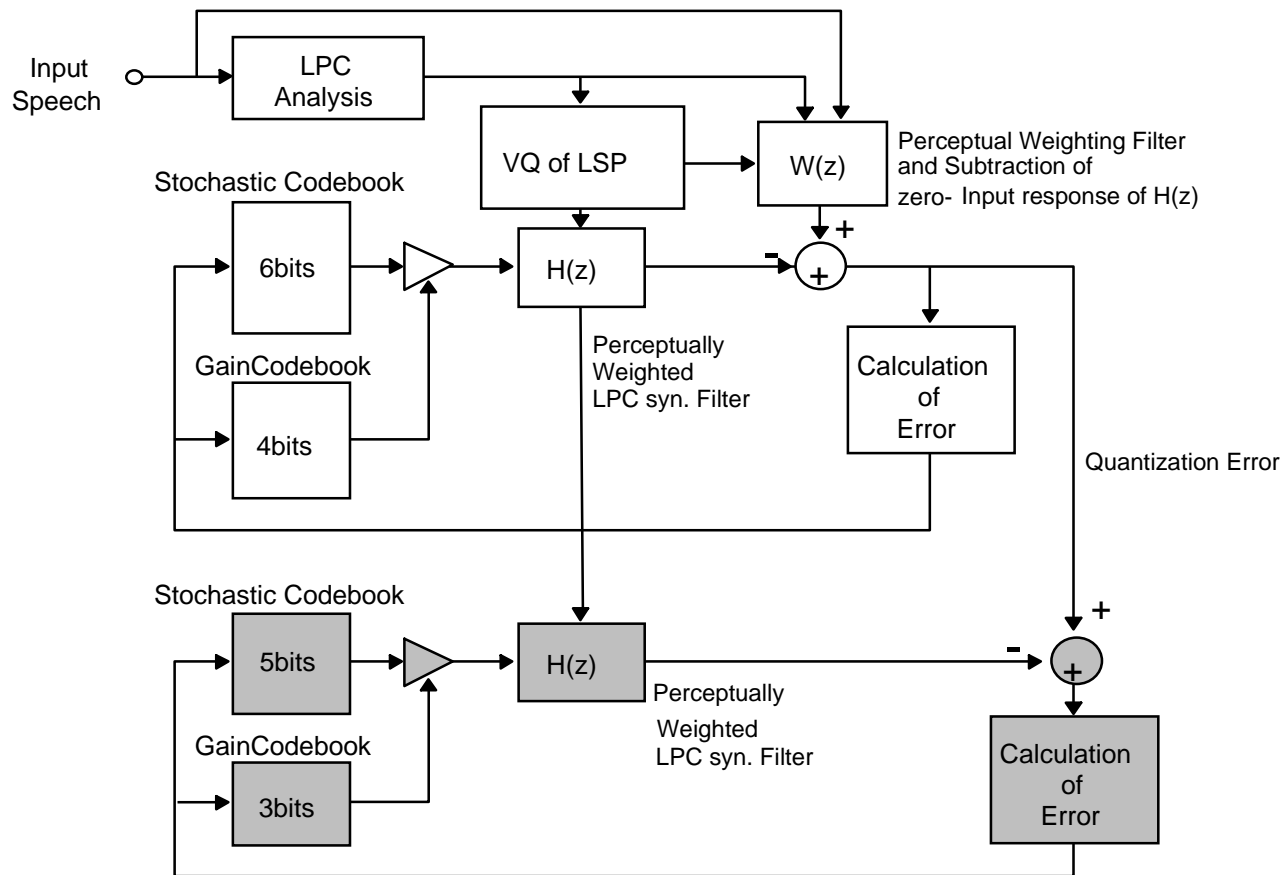
Scalable vector quantization of spectral envelope - base & enhancement layer -



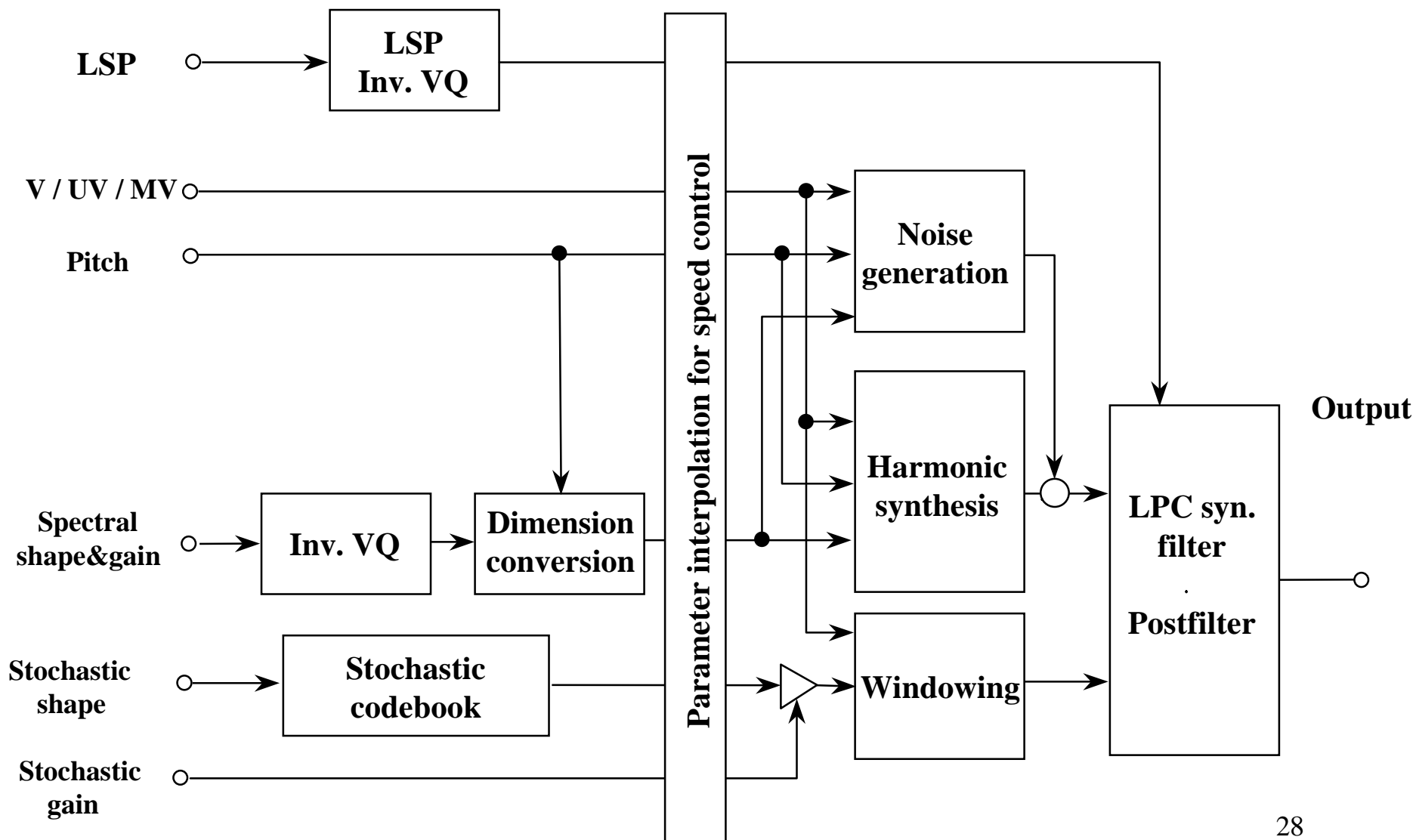
Encoder



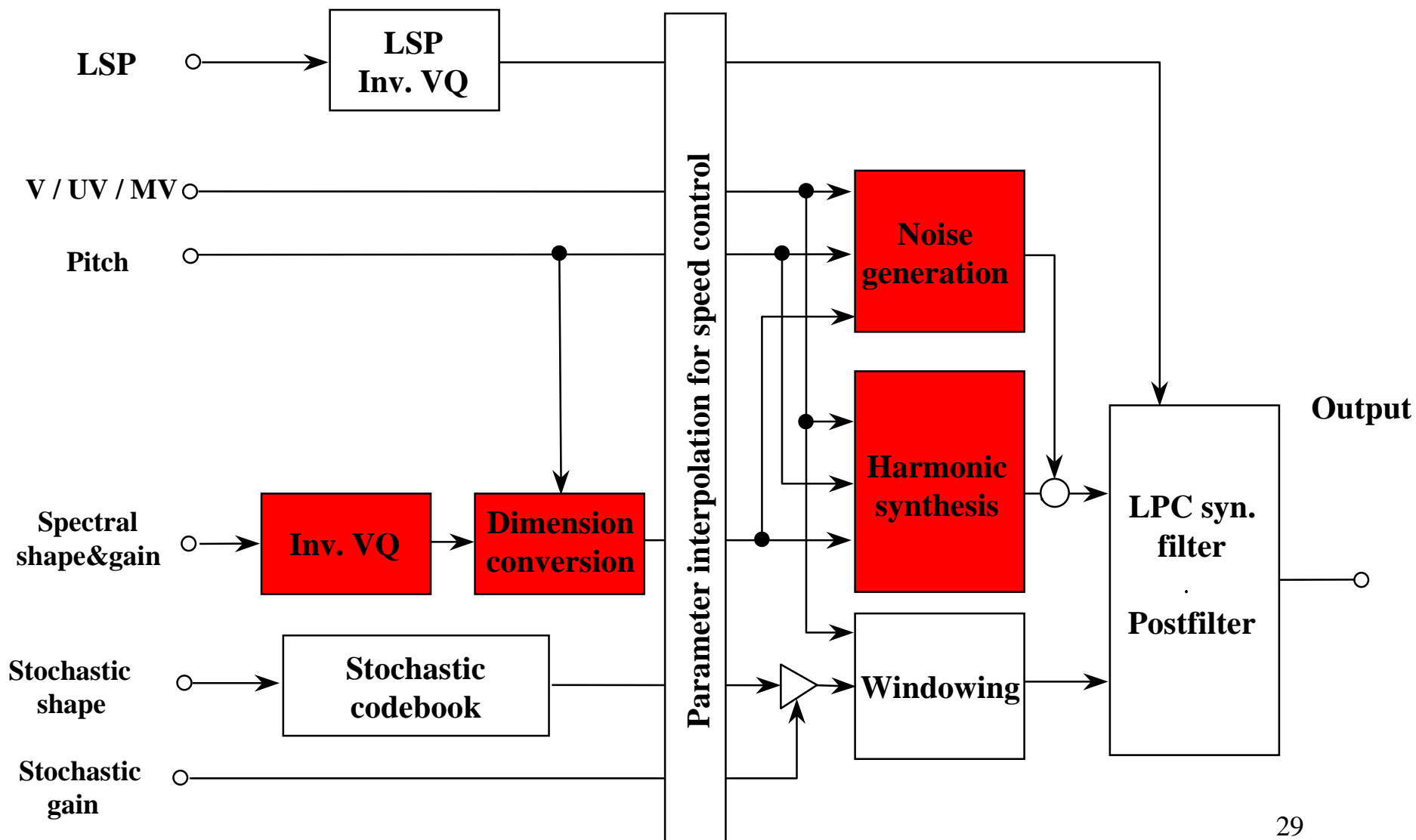
Scalable CELP encoder for unvoiced segments - base and enhancement layer -



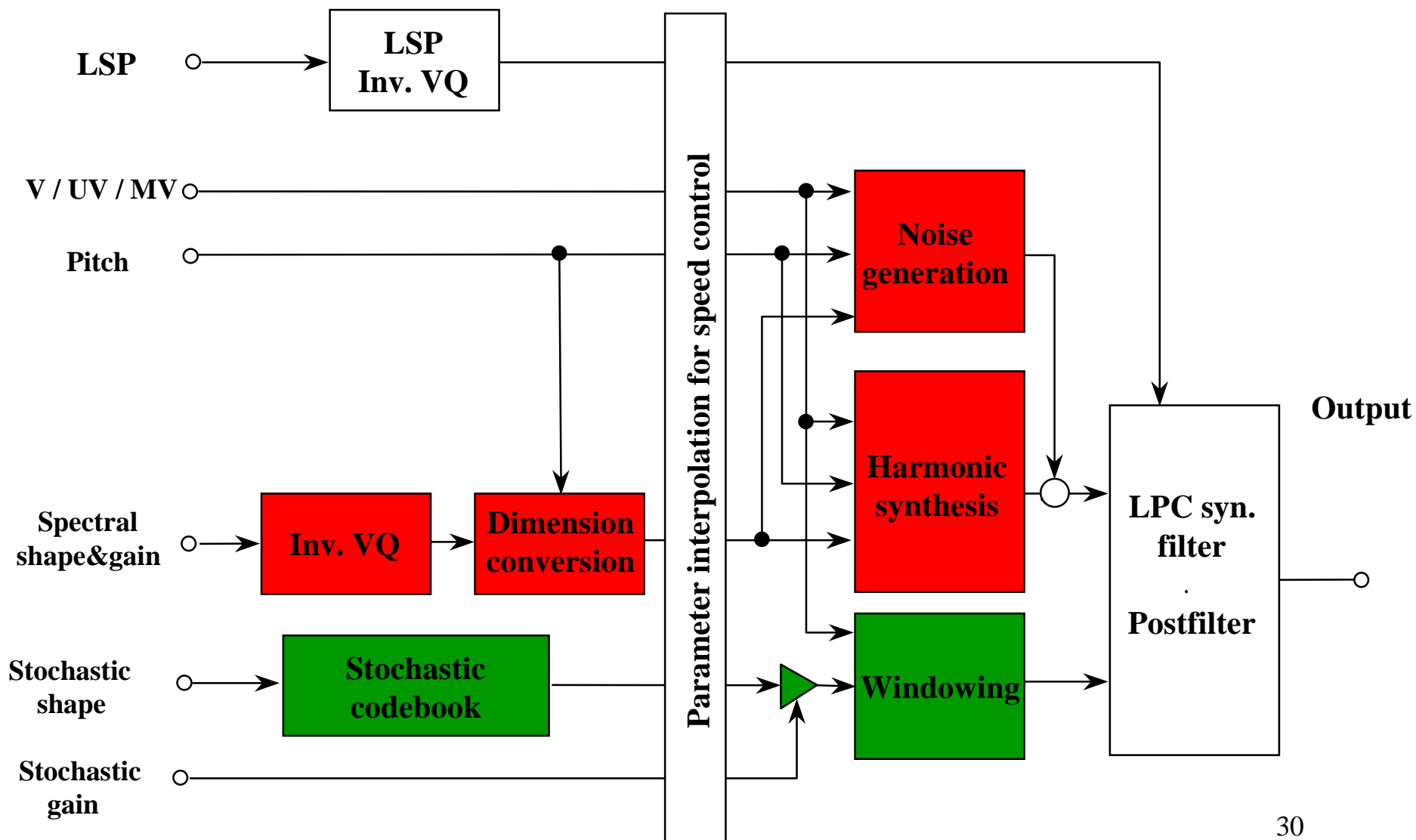
Decoder



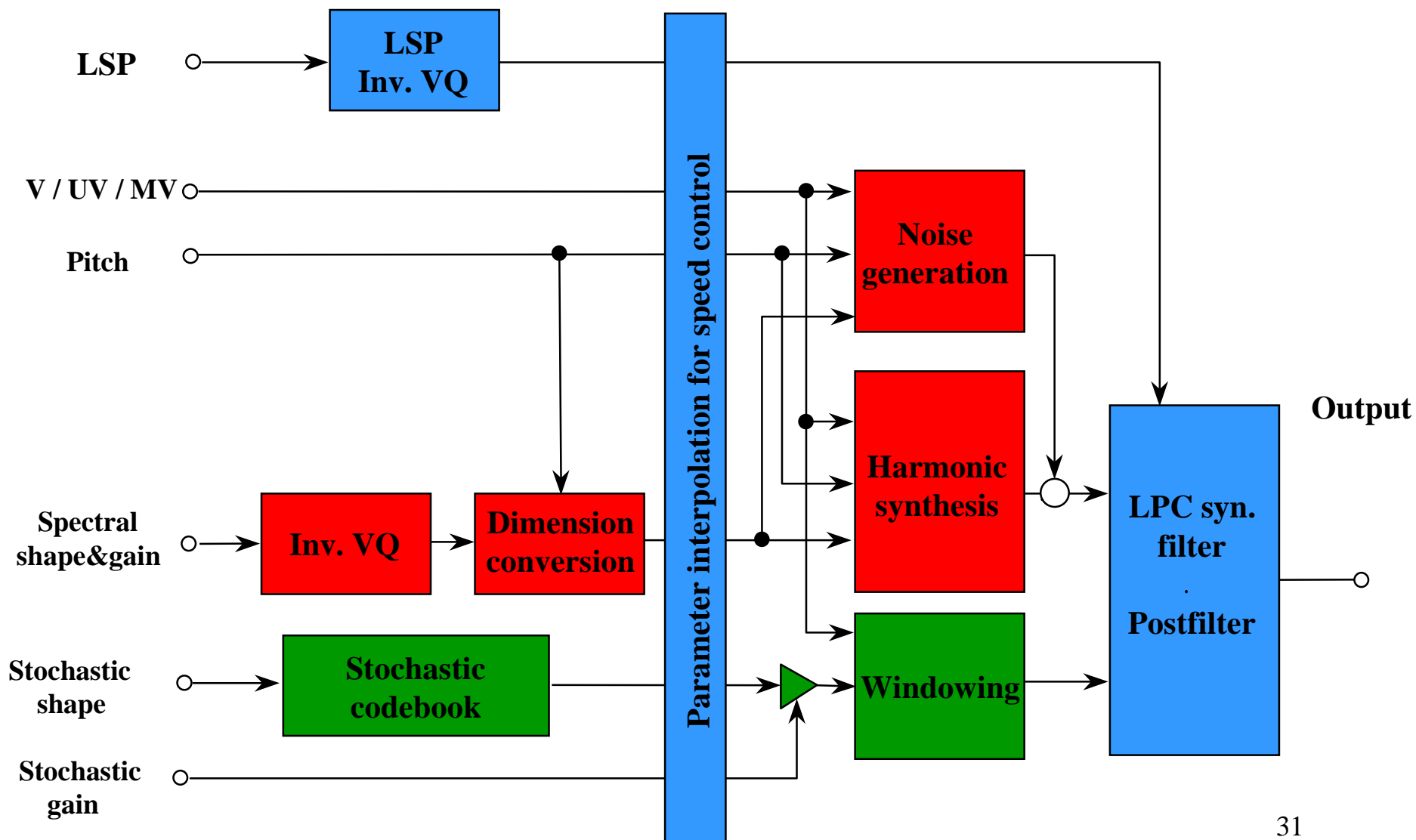
Decoder



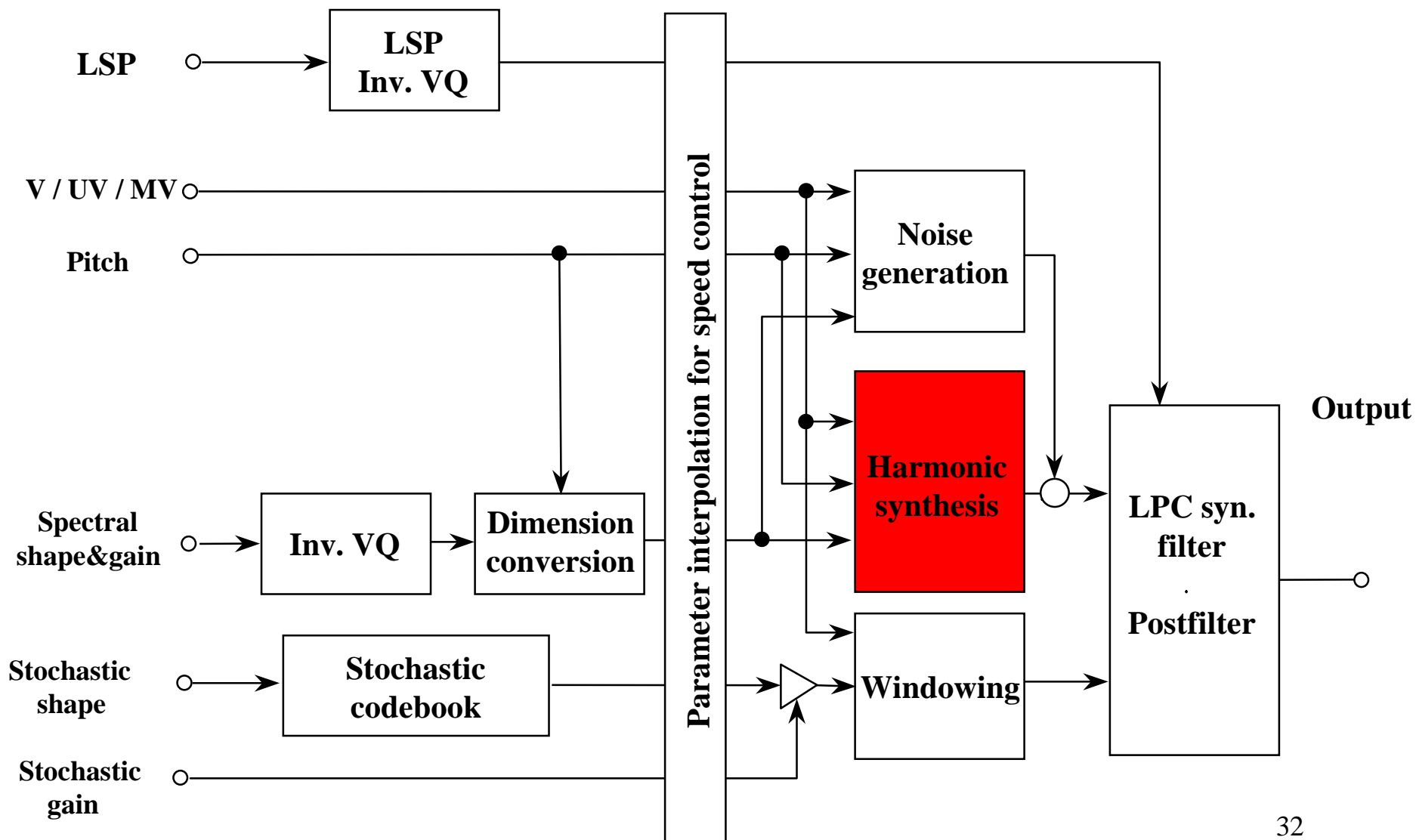
Decoder



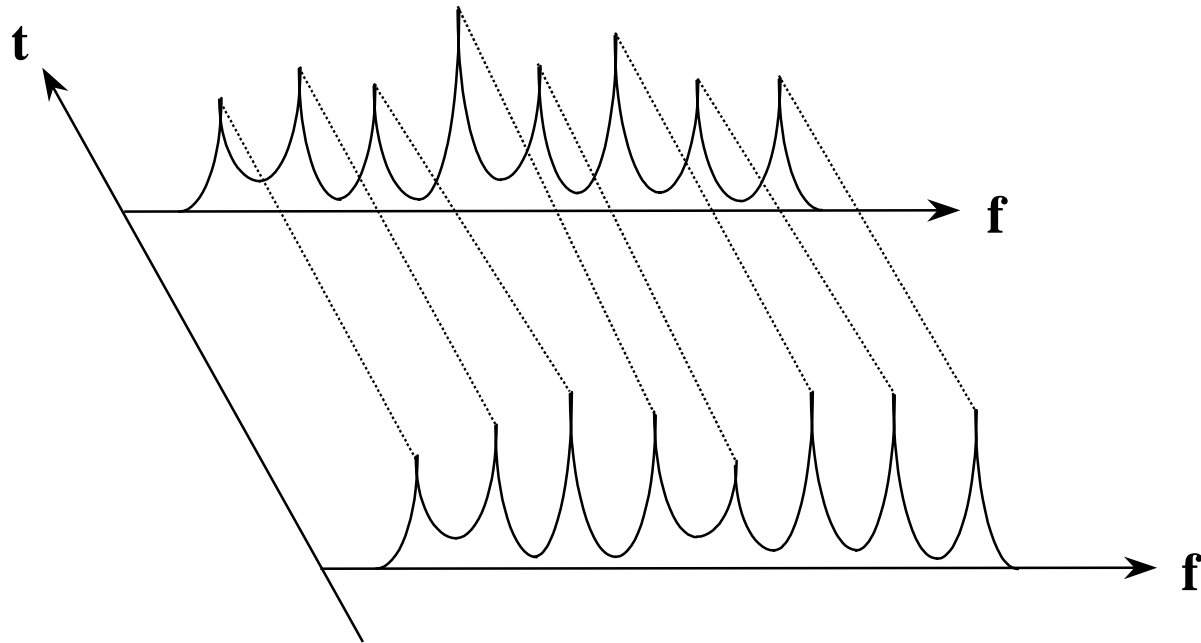
Decoder



Decoder



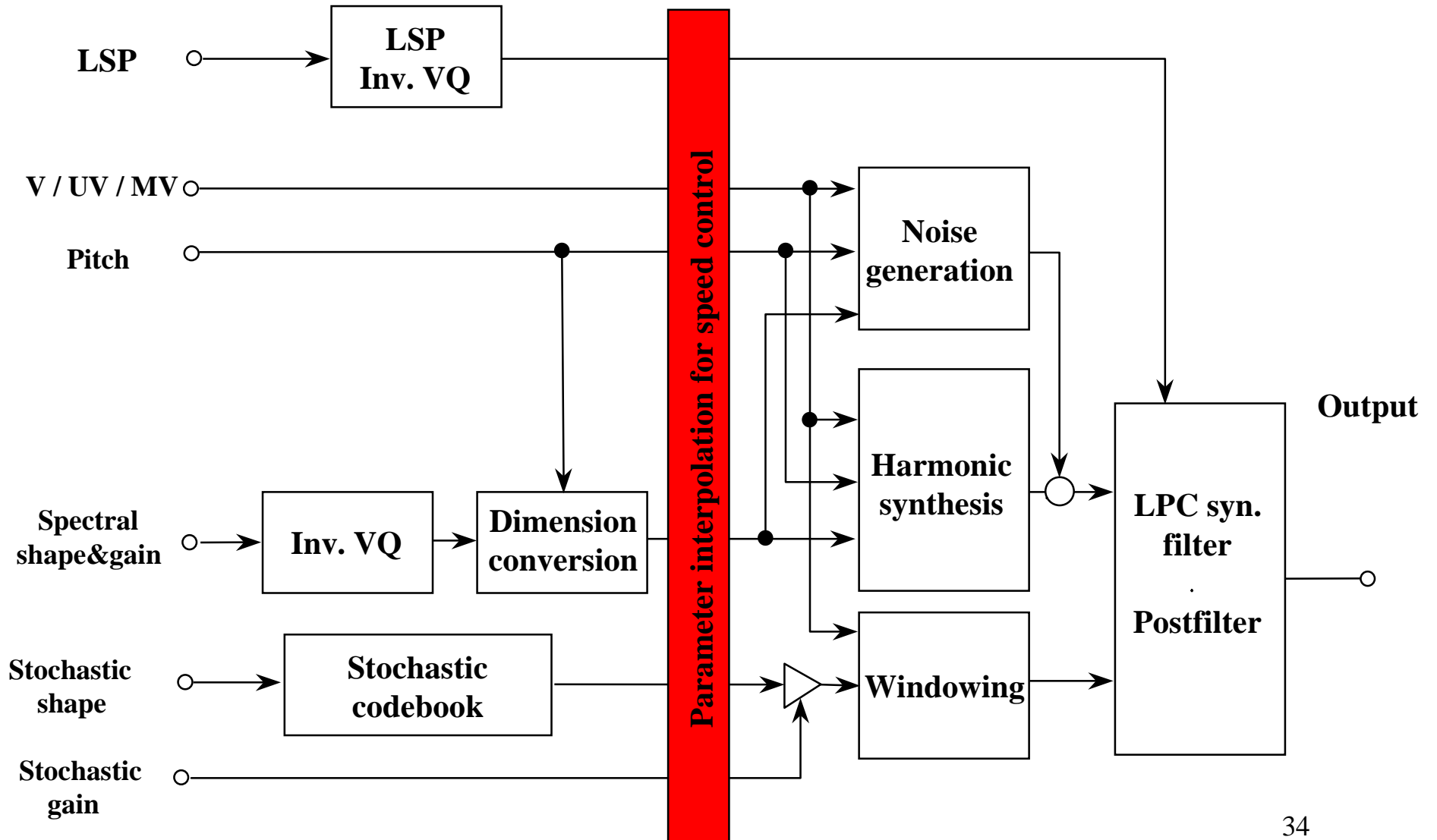
Harmonic synthesis for voiced excitation



$$f(t) = \sum_m A_m(t) \cos(\theta_m(t))$$

$$\theta_m(t) = \int_0^t \omega_m(\tau) d\tau + \phi_0$$

Decoder



Parameter interpolation for speed control

arrays of original parameters : $param [n]$

arrays of interpolated parameters : $mdf_param [m]$

time index before the time scale modification : n

time index after the time scale modification : m

ratio of speed change : spd $spd > 1$ speed up
 $spd < 1$ speed down

define:

$$\begin{cases} fr_0 = \lceil m * spd \rceil - 1 \\ fr_1 = fr_0 + 1 \end{cases}$$

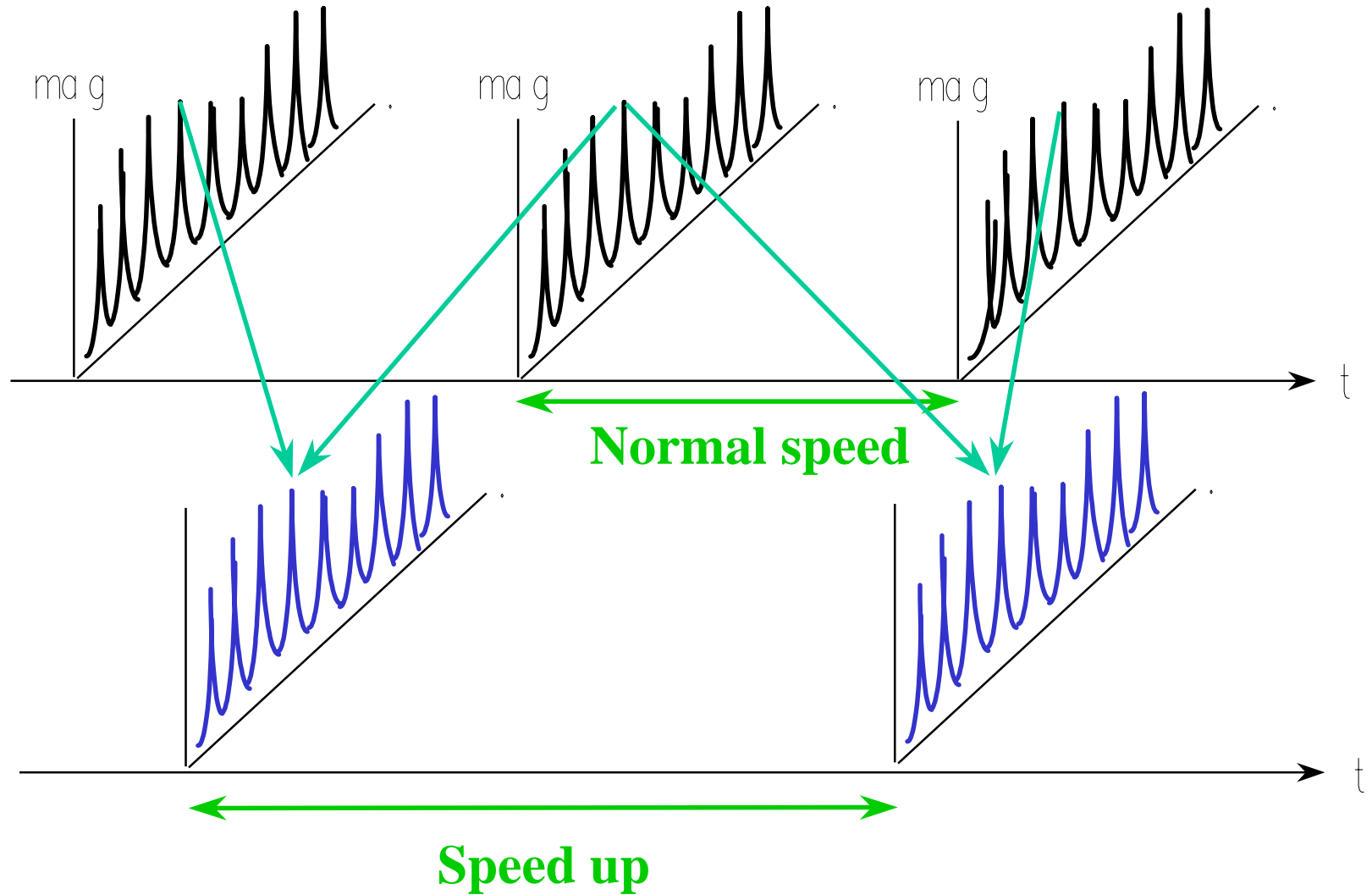
define:

$$\begin{cases} l = m * spd - fr_0 \\ r = fr_1 - m * spd \end{cases}$$

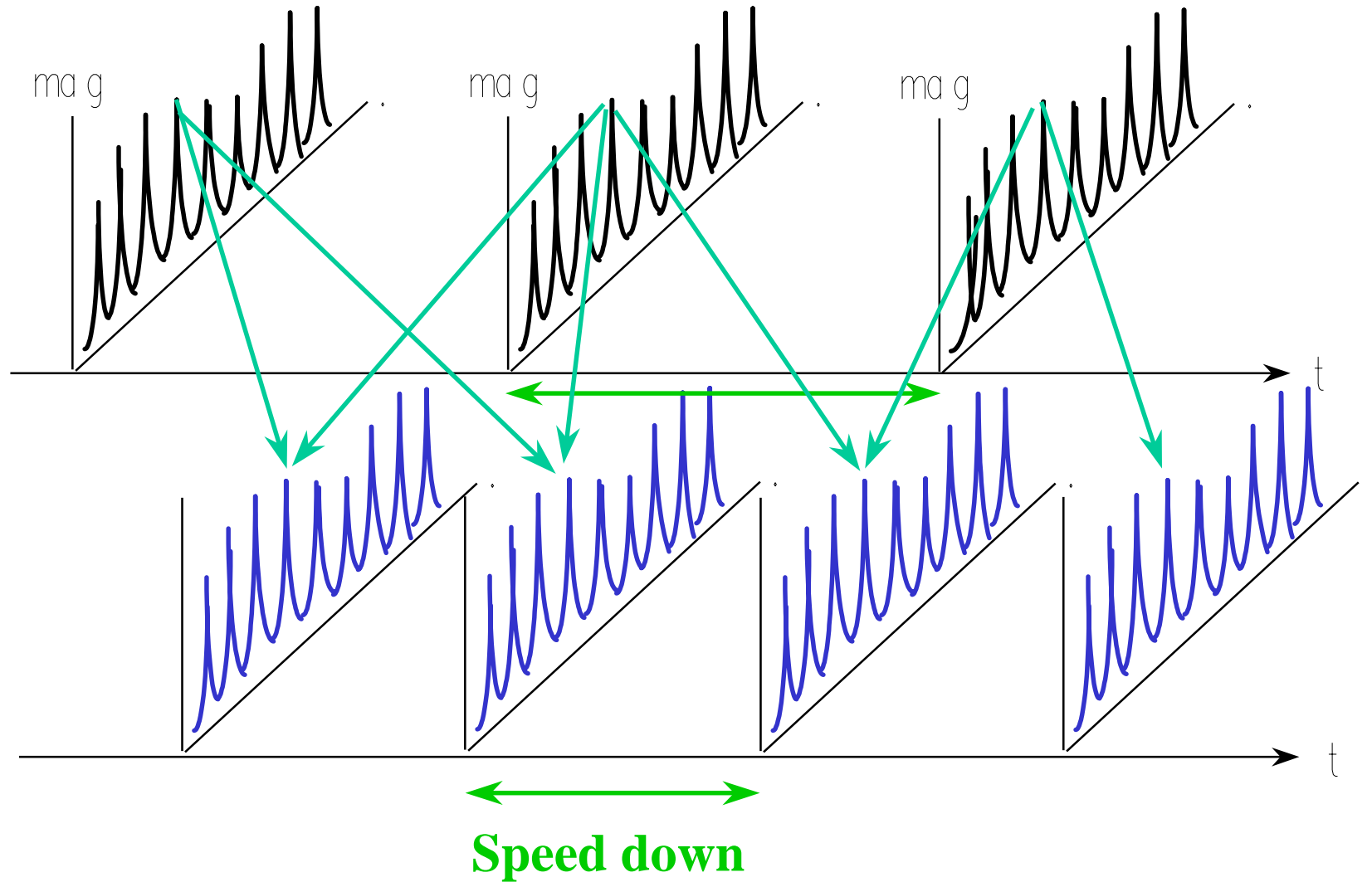
time scale modified parameters are approximated as:

$$mdf_param[m] = param [fr_0] * r + param [fr_1] * l$$

Harmonics spectra interpolation for speed control



Harmonics spectra interpolation for speed control



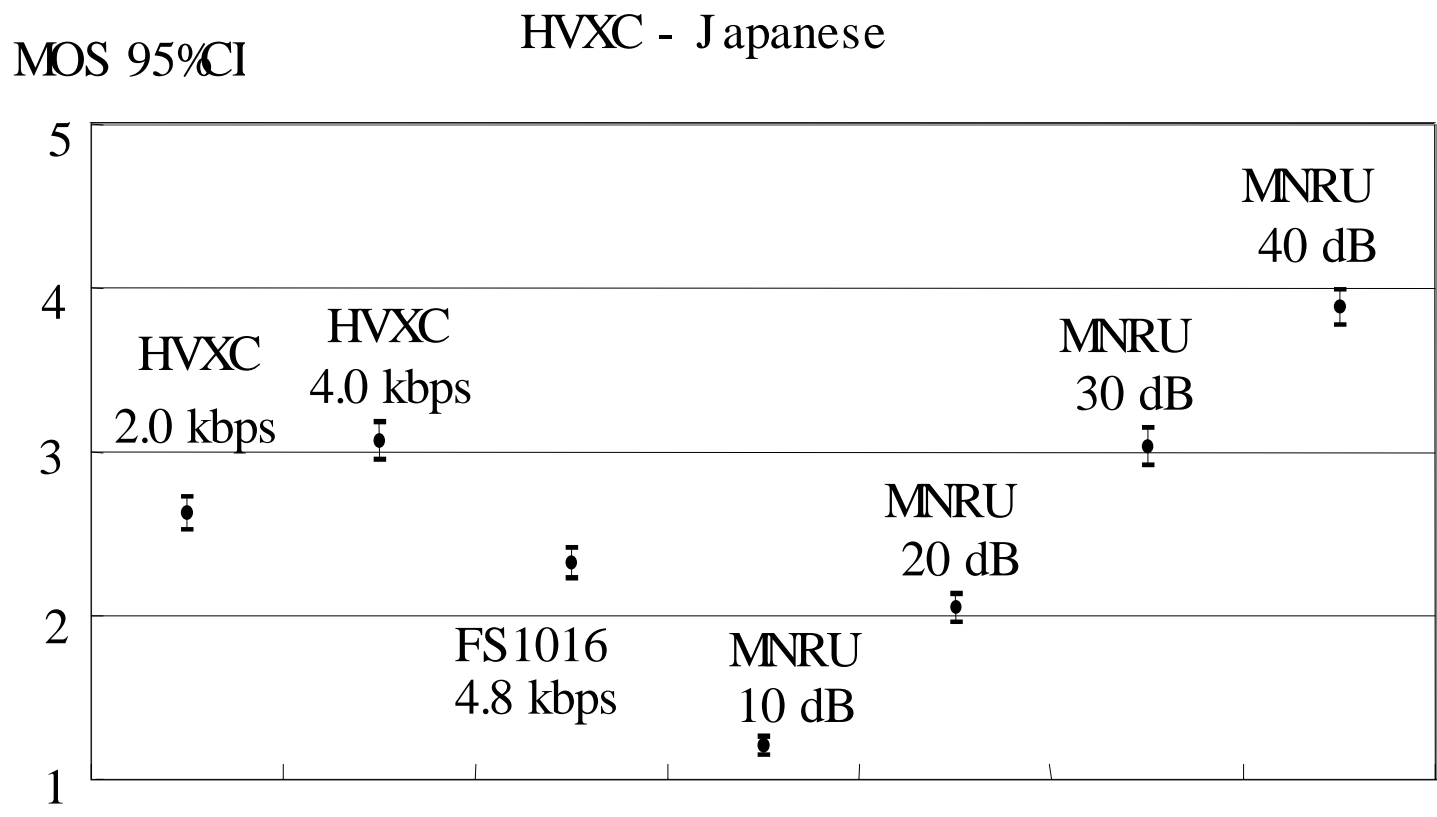
Performance

Speech quality of MPEG-4 speech codecs were evaluated in the official MPEG-4 verification tests in Aug 98 at 2 European labs and 1 Japanese lab*.

15 Japanese items were evaluated by 16 Japanese listeners in the Japanese Lab.

15 English/German/Swedish items were evaluated by 18 German and 16 Finnish listeners in the European Labs.

* ISO/IEC JTC1/SC29/WG11 MPEG98/N2424 "Report on the MPEG-4 speech codec verification tests," Oct. 1998



ISO/IEC JTC1/SC29/WG11 MPEG98/N2424 "Report on the MPEG-4 speech codec verification tests," Oct. 1998

MPEG-4 HVXC Demonstration

FS1016 4.8kbps CELP

2kbps HVXC

4kbps HVXC

Demonstration

Real time software decode by PC

4kbps HVXC pitch change

4kbps HVXC speed change

Summary

- HVXC at 2.0kbps and 4.0kbps > FS1016 CELP at 4.8 kbps.
- NB CELP \approx existing standards at the same bit-rate ranges providing flexible bit-rate controllability and scalability.
- WB CELP at 18kbps \approx G.722 at 48 to 56 kbps.
- MPEG-4 speech coding provides new functionalities
 - speed and pitch change
 - bit-rate / bandwidth scalability
 - bit-rate controllability
- International Standard in November 1999

References

- [1]ISO/IEC JTC1/SC29/WG11 N2503, "Final Draft International Standard of ISO/IEC 14496-3", Dec. 1998
- [2]M.Nishiguchi, K.Iijima, J.Matsumoto, "Harmonic Vector Excitation Coding of Speech at 2.0 kbps," IEEE Workshop on Speech Coding, Sep.1997
- [3]T.Nomura, M.Iwadare, M.Serizawa, K.Ozawa, "A Bit rate and Bandwidth Scalable CELP coder," Proc. ICASSP-98, pp.I-341-344, May. 1998
- [4]T.Nomura, M.Iwadare, N.Tanaka,"MPEG-4/CELP speech coding Algorithm," Tech. Report of IEICE, SP98-89, Nov. 1998
- [5]M.Nishiguchi, A.Inoue, Y.Maeda, J.Matsumoto," Parametric Speech Coding – HVXC at 2.0-4.0 kbps,"IEEE Workshop on Speech Coding, June 1999
- [6]N.Tanaka, et al.,"A Multi-mode Variable Rate Speech Coder for CDMA Cellular Systems",
Proc. IEEE VTC pp.198-202, Apr.1996
- [7]D.W.Griffin and J.S.Lim, "Multiband Excitation Vocoder,"
IEEE Trans. ASSP, Vol.36, pp.1223-1235, Aug. 1988
- [8]M.Nishiguchi, J.Matsumoto, S.Ono, R.Wakatsuki, "Vector Quantized MBE with Simplified V/UV Division at 3.0Kbps," Proc. ICASSP-93, pp.II-151-154, Apr.1993
- [9]M.Nishiguchi, J.Matsumoto, "Harmonic and Noise Coding of LPC Residuals with Classified Vector Quantization," Proc. ICASSP-95, pp.I-484-487, May 1995
- [10]M.Nishiguchi, K.Iijima, J.Matsumoto, "Low bit rate speech coding by Harmonic Vector Excitation Coding," Proc .ASJ 1-2-4, Sep 1997
- [11]ISO / IEC JTC1 / SC29 / WG11 MPEG98 / N2424 "Report on the MPEG-4 speech codec verification tests," Oct. 1998

END